Multi-Objective Optimization of FEA Problems with Q-Learning and Design of Experiments

Pedro V. Marcal*
M pact Corp., Oak Park, Ca., USA
pedrovmarcal@gmail.com

Jeffrey T. Fong**
National Institute of Standards & Technology
100 Bureau Drive
Gaithersburg, MD 20899
fong@nist.gov

Robert Rainsberger
XYZ Scientific Applications, Inc.,
2255 Morello Avenue, Suite 220
Pleasant Hill, CA 94523
rains@truegrid.com

Introduction

Most of the problems of structural design have been cast as optimization problems subject to direct, [Galletly & Marcal,1] or implied [Prager & Marcal,2] constraints. A given structural configuration is given, a series of one or more loads are specified and this is coupled with constraints on the properties of material or displacement of the structure. In this paper, we wish to address multi-objective problems of design where the structural loading as well as the structural behavior may change in time.

In recent years Q-learning with delayed rewards[Watkins,3] has been applied to model free problems in A.I. When combined with deep learning and cast as a competitive game, it has produced spectacular results in beating the World’s Go Champion [Hassabis et al., AlphaGo,4]. The Q-Learning process in its discrete time-step form is applied to Markov processes, ie it has no memory of what has happened in the past, so that if a reward function R is defined together with a learning function gamma ( 0.0 <gamma<1.0) as a function of increments in time, then the total reward Q is given by

Q=R(t0) + gamma*R(t1) + gamma**2*R(t2) +…  (1)

To complete the specification of the problem, we define a set of States S which can be altered by a set of actions A. Each Reward R(t) is transformed to the Reward R(t+1) by the actions A at time t. At each time t we adopt an optimal policy where we optimize the value of Q going forward in time with respect to the actions A.

*Corresponding Author, **Contribution of NIST, not subject to Copyright.
By adopting an optimal policy consistently, we satisfy the requirements of Dynamic Programming [Bellman,5], which results in an optimal value of the Q_learning response. Watkins et al[6] showed that under certain conditions of action repeatability, iteration of the Q_learning policies result in convergence of the procedure in model free environments.

When we adapt Q_learning for a simulated environment, our task is simplified because we can evaluate our rewards at any point in time by say a Finite Element Analysis. To make things specific, we assume that we are dealing with linear elastic small displacement problems of beams and curved shells. Our design objective is a combination of total volume, maximum equivalent stress and maximum displacement in the structure. Because we are dealing with a structure in service and minimizing our Q_reward, we suspect that our delay function gamma should be higher than 1 to represent wear and tear in our structure.

<table>
<thead>
<tr>
<th>Object/event</th>
<th>Ln</th>
<th>Ln-1</th>
<th>L...</th>
<th>L1</th>
<th>L0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q0</td>
<td></td>
<td></td>
<td></td>
<td>R0</td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td></td>
<td></td>
<td></td>
<td>R1</td>
<td>H0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td>R1</td>
<td>H0</td>
</tr>
<tr>
<td>Qn-1</td>
<td></td>
<td>Rn-1</td>
<td>...</td>
<td>R1</td>
<td>H0</td>
</tr>
<tr>
<td>Qn</td>
<td>Rn</td>
<td>Rn-1</td>
<td>...</td>
<td>R1</td>
<td>H0</td>
</tr>
</tbody>
</table>

Table 1. Delayed Reward Q learning

Table 1 gives a pictorial form of Watkins theory given by eq. (1)

Where L is the action and the l.c. indicates the event time n

- Gamma is the time reward function
- R is the Objective function
- Q is the total reward at time n

Because Qn is optimized at every step, from Bellman’s principle of dynamic programming, we have an optimized function for any time event n.

Theoretical Considerations.

In our work, we use the general purpose program MPACT [7] with its highly accurate 27-node hexa. This element only requires one element through the thickness[8]. In our analysis we use the top surface nodes as action variables A. These nodes are constrained to move along a straight line joining the top surface nodes to their equivalent bottom surface nodes. Finally we need to choose an optimization procedure. Normally we would choose a variant of the method of steepest descent. However in the current problem, because of the indirect nature of our design objective, we wish to select a method that emphasizes cross differentials. Because of our past experience with it, we chose to adopt a method of Design of Experiment (DOE) [box,9]. The method uses a two level fractional factorial set of runs (series) as described in [nist,10], where it is emphasized that the selection from a full factorial must be balanced and orthogonal. In order to obtain rates of change, we use all the nodes of an external surface as factors in our experiment. We allow each factor in our design to take +/- 0.1 * thickness. This allows us to select the optimal change from our series of runs (minimum). This gives us the optimal rate of change, which we will refer to as the DOE gradient. To find the optimal in the direction of the gradient, we use the usual alpha search technique. We first double the increment (alpha=2), if...
the result is successful, we continue using alpha in our search. When we have a failure we search again with a value of alpha=0.25 of the failed increment and terminate our search.

The time function of our problem is governed by the incremental loads that we apply to our structure. Because of the general nature of our program, we have the flexibility to simulate a wide range of problems. In order to make our problem specific, we will consider an initial uniform vertical load. Subsequent loading will take place over four increments of time. Each of the subsequent three loads will be 0.25 of the initial load. Each load will be applied linearly in time so that it will hit its peak in the first three time steps and then be removed in the following time step, respectively.

It’s instructive to consider the number of analysis required to solve this problem. After each loading, we need to calculate the DOE gradient using N runs. There are four loadings (not including the last). The search after each gradient is found requires an average of three additional runs. That means we need about 4*(N+3) total analysis.

Our design objective is a combination of total volume, max. equivalent stress and max. displacement, respectively. Each of the values is normalized and then summed. In practice, each of these properties are influenced differently w.r.t. the change in variables so that they should be weighted relative to each other. By trial and error it was found that optimal control of the quantities was obtained when the volume was scaled by 10. This indicates that the volume was not as sensitive as the other two quantities w.r.t. changes in the variables.

Case Study: Built-in Beam with Uniform Vertical Load

We study a built-in Steel beam with dimensions 100 X 20 X 10.5 mm. This beam is shown in its deflected position at the end of its loading in Fig. 1. The plot is of its equivalent Stress. The white frame shows its undeformed shape.
As mentioned earlier, our first series of run was to determine the best weighting between total volume and max. equivalent stress and max. displacement. The results for weights of 1 and 10 are shown in Fig. 2 and Fig. 3, respectively (gamma=0.5).
Comparing the max. stress values for the end nodes, we see a lower value for the relative weight of 10. We choose this weight for our further work.

The next results are for gamma values of 0.75 and 1.25 respectively. We would be using the first value (1.0-0.25) for a maximizing delayed reward system, whereas the (1.0+0.25) reflects a minimizing delayed reward system where the gamma value represents a further penalty due to wear and tear (e.g. abrasion, corrosion fatigue and etc.).
Fig. 5. Optimal design for $\gamma = 1.25$, showing max. equiv. stress vs Time.

Fig. 6. Optimal design for $\gamma = 0.75$, showing max. z disp. Vs Time.

Fig. 7. Optimal design for $\gamma = 1.25$, showing max. z disp. Vs Time.
We also compare the max. z disp. at the tip of the beam in Figs. 6-7. We can see that both the sig. equiv. stress values and max. z disp. Values are lower for the case where \( \gamma = 1.25 \).

Table 1. shows the optimized Q-reward and total volume after each increment of analysis by the finite element program MPACT.

<table>
<thead>
<tr>
<th>Time step</th>
<th>Q-reward</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>init</td>
<td>3.16</td>
<td>10504</td>
</tr>
<tr>
<td>0</td>
<td>1.79</td>
<td>7562</td>
</tr>
<tr>
<td>1</td>
<td>2.66</td>
<td>5445</td>
</tr>
<tr>
<td>2</td>
<td>4.78</td>
<td>5009</td>
</tr>
<tr>
<td>3</td>
<td>2.77</td>
<td>5551</td>
</tr>
<tr>
<td>4</td>
<td>5.28</td>
<td>4761</td>
</tr>
</tbody>
</table>

Table 1. Optimized Q-reward and total Volume after each time step.

We note the trend of the increasing Q-reward due to the addition of the gamma scaled previous results as per eq. (1). Because of this integration of previous results in our objective function, the resulting shape is only obtained after the final time step. The deformed shape is shown below in Fig. 8. The actual changes between the original and final nodes are given in Appendix A.

Fig. 8. Final Design Shape of deformed designed structure.

Case Study: Simply supported cylindrical Shell Roof

For our next design, we study the much studied cylindrical shell roof. [Amhad et al, 9]

Fig. 9 gives details of the problem.
Fig. 9. Quarter Model of Simply Supported Barrel Vault [7].
The thin shell is modeled by a 2 X 2 hexa 27 nodes elements.
We continue using the parameters for the Q-reward design, viz. vol. ratio =10 and gamma=1.25.
With the same incremental loading.
The results for the max. equivalent stress and max. x disp. Is given by Figs. 10-11 respectively.

Fig. 10. Optimal Design, Equivalent stress vs. Time step for Barrel Vault.
Table 2. shows the optimized Q-reward and total volume after each increment of analysis by the finite element program MPACT.

<table>
<thead>
<tr>
<th>Time step</th>
<th>Q-reward</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>init</td>
<td>3.16</td>
<td>183767</td>
</tr>
<tr>
<td>0</td>
<td>1.88</td>
<td>141335</td>
</tr>
<tr>
<td>1</td>
<td>2.16</td>
<td>113782</td>
</tr>
<tr>
<td>2</td>
<td>2.63</td>
<td>91367</td>
</tr>
<tr>
<td>3</td>
<td>5.17</td>
<td>91762</td>
</tr>
<tr>
<td>4</td>
<td>9.54</td>
<td>91789</td>
</tr>
</tbody>
</table>

Table 2. Optimized Q-reward and Total volume after each time step.
The design is obtained after the last increment. The design geometry is shown as Fig. 12.
Fig. 12. Design Geometry (deformed) for 2 X 2 Barrel Vault.

For comparative purposes, we also show the initial geometry in Fig. 13.
From the above two diagrams, we note that the position of max. equivalent stress has shifted from its position on the centerline to the mid-point along the unsupported edge.

Discussion of Results.

We have demonstrated in our case studies that we were able to use Q-learning to convert the incremental simulation into a multi-function design procedure applied to beams and shell. Since our procedure was based on changing the nodal geometry normal to an external surface, it is interesting to speculate that the method will carry over to solid models. The relative weighting of the components of our objective function was found to be an important factor and we recommend that the weights be found by trial and error as part of the design process. The choice of a gamma delay function greater than 1 as opposed to values less than 1 used in the literature for optimizing model free problems is consistent with the concept of increased damage through wear and tear. It is recommended that the gamma values above 1 be used for structural design.
Our results show that we have been able to reduce the Total Volume for both of our problems and yet be able to control the size of the other two objective functions. (see Tables 1-2) An examination of the tables also show an increasing Q-reward has to be minimized after each time step. This reflects the sensitivity of the design to all phases of the loading. This then achieves the main purpose of the Q-learning function.

Conclusions and Future Work.

We have applied the Q-delayed reward function in eq. (1) to the multi-function design of a beam and curved shell problem by post-processing the results of a finite element analysis. We have been able to come to the following tentative conclusions :-

1. Developed a program Q-FEM that calls on the MPACT FEM program to carry out the design process. The program is fairly general and may be applied to similar problems.
2. Geometry modifications were achieved through node movements normal to an external surface.
3. A Design of Experiment procedure of fractional factorials, named here as the DOE-gradient was successfully applied to act as an optimizing procedure. This was based on an intuition that the cross derivative terms of our objective function was important.

Much work remains to be done to explore design with a general purpose multi-physics program such as MPACT. The application of Q-FEM to solid models with and without nonlinearities can be achieved with little change in our procedures.

References


